


11/21/00
JC945 U.S. PTO

Please type a plus sign (+) inside this box 

11-24-00

PTO/SB/05 (11-00)

Approved for use through 10/31/2002. OMB 0651-0032

U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 C.F.R. 1.53(b))

Attorney Docket No.	1770-3-3
First Inventor	TRIPP, Gary W.
Title	INDEX CARDS ON NETWORK HOSTS FOR SEARCHING, RATING, AND RANKING
Express Mail Label No.	EJ754983845US

JC953 U.S. PTO
09/718531
11/21/00

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

- ☒ Fee Transmittal Form (e.g., PTO/SB/17)
(Submit an original and a duplicate for fee processing)
- ☒ Applicant claims small entity status.
See 37 CFR 1.27.
- ☒ Specification [Total Pages 43]
(preferred arrangement set forth below)
 - Descriptive title of the Invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to sequence listing, a table, or a computer program listing appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
- ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets 5]
- ☒ Oath or Declaration [Total Pages 3]
 - ☒ Newly executed (original or copy)
 - ☐ Copy from a prior application (37 CFR 1.63 (d))
(for a continuation/divisional with Box 18 completed)
 - ☐ DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).
- ☐ Application Data Sheet. See 37 CFR 1.76

ADDRESS TO:

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

ACCOMPANYING APPLICATIONS PARTS

- ☒ Assignment Papers (cover sheet & document(s))
- ☐ 37 C.F.R. §3.73(b) Statement (when there is an assignee) ☒ Power of Attorney
- ☐ English Translation Document (if applicable)
- ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
- ☐ Preliminary Amendment
- ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)
- ☐ Certified Copy of Priority Document(s)
(if foreign priority is claimed)
- ☐ Request and Certification under 35 U.S.C. 122 (b)(2)(B)(i). Applicant must attach form PTO/SB/35 or its equivalent.
- ☐ Other: _____

18. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment, or in an Application Data Sheet under 37 CFR 1.76:

☐ Continuation ☐ Divisional ☒ Continuation-in-part (CIP)
Prior application information: Examiner _____

of prior application No 09 1419405
Group / Art Unit. _____

For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 5b, is considered a part of the disclosure of the accompanying or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

17. CORRESPONDENCE ADDRESS

☒ Customer Number or Bar Code Label

000996

or ☐ Correspondence address below

(Insert Customer No. or Attach bar code label here)

Name	Jeffrey T. Haley				
	Graybeal Jackson Haley LLP				
Address	155 - 108th Avenue Northeast Suite 350				
City	Bellevue	State	WA	Zip Code	98004-5901
Country	USA	Telephone	425-455-5575	Fax	425-455-1046

Name (Print/Type)	Jeffrey T. Haley	Registration No. (Attorney/Agent)	34,834
Signature		Date	November 21, 2000

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**FEE TRANSMITTAL
for FY 2001**

Patent fees are subject to annual revision.

Complete if Known

TOTAL AMOUNT OF PAYMENT (\$) 1259

Application Number

Filing Date

First Named Inventor

TRIPP, Gary W.

Examiner Name

Group / Art Unit

Attorney Docket No.

1770-3-3

METHOD OF PAYMENT (check one)1. ☒

The Commissioner is hereby authorized to charge indicated fees and credit any over payments to:

Deposit
Account
Number

07-1897

Deposit
Account
Name

Graybeal Jackson Haley LLP

☐ Charge Any Additional Fee Required
Under 37 CFR 1.16 and 1.17☒ Applicant claims small entity status.
See 37 CFR 1.272. ☒ Payment Enclosed:☒ Check ☐ Credit card ☐ Money
Order ☐ Other**FEE CALCULATION****1. BASIC FILING FEE**

Large Fee Code	Entity Fee (\$)	Small Fee Code	Entity Fee (\$)	Fee Description	Fee Paid
101	710	201	355	Utility filing fee	355
106	320	206	160	Design filing fee	
107	490	207	245	Plant filing fee	
108	710	208	355	Reissue filing fee	
114	150	214	75	Provisional filing fee	

SUBTOTAL (1)

(\$) 355

2. EXTRA CLAIM FEES

Total Claims	Extra Claims	Fee from below	Fee Paid
76	-20** = 56	X 9 = 504	
Independent Claims	12	-3** = 9	X 40 = 360
Multiple Dependent		X	= 0

Large Fee Code	Entity Fee (\$)	Small Fee Code	Entity Fee (\$)	Fee Description
103	18	203	9	Claims in excess of 20
102	80	202	40	Independent claims in excess of 3
104	270	204	135	Multiple dependent claim, if not paid
109	80	209	40	** Reissue independent claims over original patent
110	18	210	9	** Reissue claims in excess of 20 and over original patent

SUBTOTAL (2)

(\$) 864

**or number previously paid, if greater. For Reissues, see above

FEE CALCULATION (continued)

Fee Code	Large Entity Fee (\$)	Small Entity Fee Code	Small Entity Fee (\$)	Fee Description	Fee Paid
105	130	205	65	Surcharge - late filing fee or oath	
127	50	227	25	Surcharge - late provisional filing fee or cover sheet.	
139	130	139	130	Non-English specification	
147	2,520	147	2,520	For filing a request for reexamination	
112	920*	112	920*	Requesting publication of SIR prior to Examiner action	
113	1,840*	113	1,840*	Requesting publication of SIR after Examiner action	
115	110	215	55	Extension for reply within first month	
116	390	216	195	Extension for reply within second month	
117	890	217	445	Extension for reply within third month	
118	1,390	218	695	Extension for reply within fourth month	
128	1,890	228	945	Extension for reply within fifth month	
119	310	219	155	Notice of Appeal	
120	310	220	155	Filing a brief in support of an appeal	
121	270	221	135	Request for oral hearing	
138	1,510	138	1,510	Petition to institute a public use proceeding	
140	110	240	55	Petition to revive - unavoidable	
141	1,240	241	620	Petition to revive - unintentional	
142	1,240	242	620	Utility issue fee (or reissue)	
143	440	243	220	Design issue fee	
144	600	244	300	Plant issue fee	
122	130	122	130	Petitions to the Commissioner	
123	130	123	130	Petitions related to provisional applications	
126	180	126	180	Submission of Information Disclosure Stmt	
581	40	581	40	Recording each patent assignment per property (times number of properties)	40
146	710	246	355	Filing a submission after final rejection (37 CFR § 1.129(a))	
149	710	249	355	For each additional invention to be examined (37 CFR § 1.129(b))	
179	710	279	355	Request for Continued Examination (RCE)	
169	900	169	900	Request for expedited examination of a design application	

Other fee (specify) _____

*Reduced by Basic Filing Fee Paid

SUBTOTAL (3)

(\$) 40

SUBMITTED BY

Complete (if applicable)

Name (Print/Type)

Jeffrey T. Haley

Registration No. Attorney/Agent

34,834

Telephone

425-455-5575

Signature

Date

November 22, 2000

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231

PATENT APPLICATION - CERTIFICATE OF MAILING

Inventors: TRIPP, Gary W.
 MEADWAY, Michael D.
 DUGUAY, Claude E.

Title of Invention: INDEX CARDS ON NETWORK HOSTS FOR SEARCHING,
RATING, AND RANKING

Attorney Dkt. No.: 1770-3-3

jc853 U.S. PTO
09/718531
11/21/00

CERTIFICATE OF MAILING

"Express Mail" mailing label number: EJ754983845US

Date of Deposit: November 21, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR, Section 1.10 on the date indicated above and is addressed to Box New Application, Assistant Commissioner for Patents, Washington, D.C. 20231 by



Signature

Enclosures:

Utility Application For Patent Cover Sheet
Specification (43 pages)
Formal Drawings (5 PAGES)
Fee Transmittal for FY 2001
Check No. 16171 in the amount of \$1,219.00 for filing fee
Assignment of U.S. and Foreign Rights in Invention
Recordation Cover Sheet
Check No. 16172 for \$40.00 filing fee
Return postcard

INDEX CARDS ON NETWORK HOSTS FOR SEARCHING, RATING, AND RANKING

TECHNICAL FIELD

The present invention relates to computer network search engines and more particularly to a method and system for allowing the managers of host computers and of web
5 sites on those hosts to enter data into a data structure which can be used by a central search engine for searching, ranking search results, and rating web sites.

BACKGROUND OF THE INVENTION

In the last several years, the Internet has experienced exponential growth in the number of web sites and corresponding web pages contained on the Internet. Countless
10 individuals and corporations have established web sites to market products, promote their firms, provide information on a specific topic, or merely provide access to the family's latest photographs for friends and relatives. As a result of the rapid growth in web sites on the Internet, it has become increasingly difficult to locate pertinent information.

Search engines, such as Inktomi, Excite, Lycos, Infoseek, or FAST, are typically
15 utilized to locate information on the Internet. Upon inquiry from a user, the search engine software searches the millions of records contained in a central index. The search engine software finds matches to the search query and may rank them in terms of relevance according to predefined ranking algorithms. While most search engines accept submissions of sites for indexing, even upon such a submission, the site may not be indexed in a timely manner, if at all.

An inherent shortcoming of the method of indexing utilized in the conventional
20 search engine is that only documents stored with a mark-up language such as SGML, HTML or XML is utilized in generating the central index. Due to the format of a mark-up language web page, certain types of information may not be placed in the mark-up language tags. For example, conceptual information such as the intended audience's demographics and geographic
25 information may not be placed in an assigned tag in the document. Such information would be extremely helpful in generating a more useful index. For example, a person might want to search in a specific geographical area, or within a certain industry. Assume a person is searching for a

red barn manufacturer in a specific geographic area. Because mark-up language pages have no standard tags for identifying industry type or geographical area, the spider on the server in the conventional search engine does not have such information to utilize in generating the central index. As a result, the conventional search engine would typically list not only manufacturers
5 but would also list the location of picturesque red barns in New England that are of no interest to the searcher.

Some Internet search engines, such as Infoseek, have proposed a distributed search engine approach to assist their spidering programs in finding and indexing new web pages. Infoseek has proposed that each web site on the Internet create a local file named
10 "robots1.txt" containing a list of all files on the web site that have been modified within the last twenty-four hours. Files that have not been modified will not be indexed, saving bandwidth on the Internet otherwise consumed by the spidering program and thus increasing the efficiency of the spidering program.

There is a need for a method and system of indexing or cataloging remotely stored
15 data that allows conceptual information and other human generated information about web sites to be utilized in generating the index to allow sites to be found in a search, to make search results more meaningful, and to allow sites to be more accurately rated.

Full-text search and indexing systems such as web search engines typically have two distinct means of organizing the presentation of documents. The first means is usually a
20 categorization system (hierarchical or otherwise) which presents the documents in groups or "clusters" related by topic, content or origin. The second is dynamically generated as a result of a search process of some sort such as a matching keyword search. Normally, this second means is presented as a linear list in which matching documents are sorted either by alphabetical title, date of change or a ranking value based on a calculation whose input may come in part from the
25 document content. For example, in searching for the work "car" in a set of documents, the resulting list of matching documents might be sorted by the number of times the word occurred in each document.

SUMMARY OF THE INVENTION

According to one aspect of the present invention, one or more brochure files or "index cards" are created and stored within each web site on a network to provide conceptual or non-keyword data about the site, such as demographics and categorization information, related to one or more parts of the web site. This conceptual information is then utilized in constructing the central catalog so that more accurate search results may be generated in response to search queries applied to the catalog. This categorization information is transmitted by an agent program on the host to the central index.

Important uses for the categorization information include the language of the web site, any geographic location to which the web site is particularly relevant, demographic information about the audience of people that will likely be interested in the site or that the owner of the site wishes to target, or the subject matter of the site so that it can be found by people who are searching for certain subject matter.

Like other information about a web site, information contained within the index cards can be used in ranking algorithms on a search engine to rank the references found according to a set of ranking rules. The ranking process can include both data derived from the contents of objects or from meta data associated with objects, or from the index cards. In addition to using the categorization information to filter search results to produce more meaningful search results, the categories can be structured as hierarchies with sub-categories in each category. Then, when a user designates a category to find web sites that match that category designation, the system can also provide to the user references to web sites associated with sub-categories within that category.

The preferred method of transferring the index card information to the central site is with an agent program located on each web site host. To further improve the ranking results, the agent program can also process objects on the web site to parse their contents and extract data from the objects which can be used by the agent program in a ranking algorithm which follows a set of ranking rules, producing assigned rankings for objects on the web site. The agent can then transmit via signed rankings to the central cataloging site for further processing.

The index cards can also be useful for improving the rating of web sites such as rating for adult content or other content that selected groups might consider offensive. The web site owner can apply a rating to the site which is stored in the index card. Also, the administrator of the host computer system can apply a rating to the web site. These ratings can be compared to ratings which are automatically generated by parsing words on the web site to assess their probable accuracy. The agent program can assist in this process by parsing the words in the web site according to rating values and then forwarding the rating values to the central cataloging computer.

The index cards can also be used by web site owners and host system administrators to regulate the exposure of objects on the host to other computer systems on the network. The index card can be used to list non-html objects and provide sufficient information about the objects that they can be remotely accessed across the network. Remote computer systems can read the index card directly to obtain its information or, in the preferred embodiment, the index card information is periodically transmitted to a central site where it is organized for searching by remote computer systems.

For purposes of the following discussion, each object consists of both contents (meaning only the essential data itself and not a header) and associated "meta data". The meta data includes all information about the contents of an object but not the contents itself. The meta data includes any information that has been extracted from the contents and is associated with the object, any header information within the object, and any file system information stored outside of the object such as directory entries. The term "object" is used only to refer to anything stored on a site of interest to a person who might access the site from the network and its associated meta data. To avoid confusion, the term "object" is not used more broadly.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram showing the architecture of a search engine for actively indexing the world wide web according to one embodiment of the present invention.

Figure 2 is a bubble chart illustrating the generation and processing of a brochure file in the indexing system of Figure 1.

Figure 3 is a bubble chart illustrating the process executed by the queue manager in queuing update entries and transferring these entries to the remote queue manager.

Figure 4 is a bubble chart illustrating the overall data flow in the search engine of Figure 1.

5 Figure 5 illustrates components of the indexing system of Figure 1 for a Java-based implementation.

DETAILED DESCRIPTION OF THE INVENTION

Figure 1 is a block diagram of an indexing system for indexing the Internet using index cards according to one embodiment of the present invention. The system includes a central
10 server 202 that stores a central index and processes search queries received over the Internet. The system also includes brochure files or "index card" files 206 residing on respective remote servers 208, each index card file containing non-HTML or conceptual information about the web site for use in generating the central index on the server 202, as will also be explained in more
15 detail below. For the sake of brevity, only two remote servers 208 and the corresponding agent 204 and index card 206 are shown in Figure 1. The system, however, includes numerous such remote servers 208 with index card files 208.

The central server 202 includes a router 210 that directs packets comprising search requests and update transactions to an appropriate set of servers 214, 302 and 222. The web servers 214 receive and preprocess index queries and receive and process index card 206
20 generation or modification requests. The central server 202 further includes a master index server 218 containing a master copy of the entire central search index or catalog. A number of update servers 222 each receive updates from each host 208.

The central search engine 202 further includes an index card database server 226 and index card check server 228. The index card database server 226 stores an index card
25 database as a list of index card files and their associated data fields for each web site. The web servers 214 may request records from or add records to this index card file database depending on the actions taken by web site administrators while maintaining their index card entries. The index card check server 228 periodically checks for valid new index card files as defined within

the index card database server for web sites that are not being processed by a local agent program, as will be described in more detail below. If the defined index card in the index card database server 226 is not found by the index card check server 228, a notification is sent to the administrator of the site where the index card was supposed to be found.

5 Index card information may be automatically sent by each host to the central system by use of an agent program 204 on the host. Alternatively, instead of using an agent program on each host to send index card information to the central system, index cards may be retrieved directly by external central system programs, such as web crawlers, and the contents used to create categorized sets of web pages. This approach requires that the web server be capable of
10 retrieving a list of the available index cards, or that the cards have predetermined (standardized) names. Once index cards have been retrieved by such systems, the content of the index card may be used to calculate rating values.

 When an index card file is requested for a site which is not served by an agent 204, a message is sent to the Internet Service Provider ("ISP") or system administrator for the
15 site hosting the web site, indicating that users of the system are requesting index cards. This server also periodically checks the validity of existing index cards on all sites and notifies the web site administrator if an index card file is missing. If an index card is missing and remains missing for a given number of check cycles, the index card check server 228 sends a request to the index card database server 226 to delete the entry for the index card. The index card check
20 server 228 detects any changes in index card files, such as additions or removals, and converts these changes to transaction batches that are forwarded to a queue manager which, in turn, applies these changes to update the central index on the master index server 218, as will be described in more detail below. The index card check server 328 periodically verifies the status of all index cards at sites that are not being indexed by an agent 204.

25 The components of the central server 202 and their general operation have been described, and now the operation of the agent 204 and index card 206 will be described in more detail. Additional details about the agent technology is disclosed in US patent application 09/575,971 which is incorporated by reference.

The agent 204 and index card 206 may both be present at a remote server 208. An index card 206 and agent can function independently of each other, as will be discussed in more detail below. The agent 204 is a small local program which executes at the remote server 208 and generates an incremental search engine update for all of the participating web sites on the web host 208. These index updates are transmitted by the agent 204 to the central server 202, where they are queued for addition to the central index.

The agent 204 runs on a system, such as a web host server, at the site of an organization, and processes content (objects) for all web sites available via mass storage from that system. The agent 204 processes all web sites located within the mass storage area to which it has access, unless configured to exclude some portion of a site or sites. The agent 204 uses the local web server configuration (object catalog or file system information) data to determine the root directory path (or other location information for the particular file system) for all web site file structures available.

While indexing the web sites at the remote server 208, the agent 204 recognizes index cards 206 stored at web sites on the server and provides index updates based on the contents of the index card files found. Once the agent 204 has indexed the web sites at the remote server 208, the agent transmits a transaction list to the central server 202, and this transaction list is stored on one of the update servers 222. The transaction list is referred to as a batch, and each batch contains a series of deletion and addition transactions formatted as commands.

The index card 206 is a small file that may contain conceptual and other non-mark-up language information which would be useful to improve the indexing of sites or parts of a single site on the remote server 208. An index card 206 may contain any information pertinent to the web site, including but not limited to keywords, phrases, categorizations of content, purpose of the site, and other information not generally stored in a web page. The index card 206 is generated manually by individual web site administrators. The administrator fills out a form at the central server 202, and receives an email containing the index card 206 or downloads the index card file after submitting the form contents. Upon receiving the index card 206, the administrator stores it within the file structure of the web site on the remote server 208. There

may be multiple index cards 206 at the same web site, each describing specific portions of the site. Each index card 206 may refer to a single web page or a group of web pages stored within a specific subdirectory at the web site. All information stored in each index card 206 is applied to the pages referenced in the index card.

5 The central server 202 also allows remote users to generate and download index cards 206 to their remote site, and also verifies the validity of index cards 206 on web sites not serviced by an agent 204, as will now be explained in more detail. The web servers 214 receive and process index card 204 generation or modification requests from user browsers. Once the index card 204 has been generated or modified, the index card is transferred to the index card
10 database server 226, which stores all existing index cards. The index card check server 228 periodically checks for new index cards 206 stored on the index card database server 226 for web sites that are not served by an agent 204. When an index card 206 is requested for a web site which is not served by an agent 204, the index card check server 228 sends a message to the system administrator or Internet service provider for the server hosting a web site telling them
15 that site administrators on their server are requesting index cards 206. The index card check server 228 also periodically verifies the validity of existing index cards 206 on all sites not serviced by an agent 204. If an index card 206 is missing for a predetermined number of verification cycles, the index card check server 228 instructs the index card database server 226 to delete the entry for that index card. The index card check server 228 also converts any
20 modifications, additions, or deletions to index cards 206 to transaction batches, and forwards these transaction batches to a queue manager 302. The queue manager 302 receives index card update transaction batches from the index card check server 228 and also receives agent update transaction batches from the agent update server 222, as will be described in more detail below.

Figure 2 is a bubble chart illustrating the generation and processing of an index
25 card 206 in the indexing system of Figure 1. As previously mentioned, the purpose of the index card 206 is to allow the web host 208 and the web site to provide specific non-HTML information to help the central server 202 index the site and to provide more relevance to query results. The index card 206 can be created in two ways. First, as part of the installation program for the agent 204, the administrator of the remote server 208 completes a form that is converted

to an encoded index card file 206, and then copied into the web directory on the remote server 208. This method of generating the index card 206 will be discussed in more detail below. The second method of generating the index card 206 utilizes an index card creator interface on the web servers 214 at the central server 202. This method will now be described in more detail
5 with reference to Figure 2.

To create an index card 206 using the index card creator interface, a user's browser 400 applies an index card generation request 402 to the associated central site web server 214. In response to the request 404, the index card creator interface generates a form which the user completes, and then sends an index card request 406 to the index card server 226,
10 which generates an encoded index card file that is then sent to the central site web server 214. The central site web server 214 then sends the encoded index card file to the user's browser 400. The encoded index card file 206 is then stored in local storage 408. Subsequent to receiving the encoded index card file 206, the user sends the encoded index card file 206 via the user's web browser 400 to the web host site storage 410 (e.g., the web site host computer).

15 The index card server 226 stores the index card data 407 in an index card database 424 on the central server 202 once it has been generated as a result of an index card generation request 404. To verify proper storage of encoded index card files 206, the index card check server 425 retrieves index card data 420 from the index card database 424 and sends a request 416 to the web host server 404 to retrieve the encoded index card file 206 from the web host site
20 storage 410. Upon successful retrieval of the index card file 206, the index card check server generates and transmits object references 422 created as a function of the index card data 420 to the queue manager 302. The queue manager 302 thereafter updates the central index to include the generated object references.

25 The directory structure of the host and web site are used to determine the relevance of the information in the index card. Information in an index card located in the root directory will apply to all sub-directories unless superceded by another index card. Information in a directory index card will apply to all subdirectories unless superceded by information in a subdirectory index card. Where an index card is placed determines the content to which the information applies. A web site owner can have as many index cards as there are pages or

directories in his site. A site owner can request that their site be excluded from the central index by checking the EXCLUDE box next to the URL and copying the index cards into the directory to be excluded.

An example of a host index card is shown in Table 1:

5 Table 1-Host Index Card Records

	Company Information:
1.	IP number
2.	Domain Name Server
3.	Type of Domain Name Server
	• HOST - Name _____
	• Company - Name _____
	• Individual - Name _____
4.	HOST name
5.	Company Name if different
6.	Contact Name
7.	Address
8.	Phone
9.	Fax
10.	Technical Contact name:
11.	Technical Contact's direct phone number
12.	Technical Contact's email address
13.	Would you like the Technical Contact to receive email notification that of every successful site index update.
14.	Business Contact name:
15.	Business Contact's direct phone number
16.	Business Contact's email address
17.	Site Languages
18.	Site Rating
19.	URL/Sites to be indexed
20.	URL/Sites to be excluded
	General Information: (optional)
1.	Area served
2.	Number of email boxes hosted _____
3.	Number of Domain Names hosted _____
4.	Number of web sites hosted. _____

The host uses the configuration section of the agent program to create site index cards. The agent program can create site index cards for an entire IP address or for any subsection of the site.

- 5 In addition to the host index card, a web site owner may also place a site index card on his web site. The purpose of the site index card is to allow the web site owner to provide specific conceptual or non-html information, which will help in indexing the site. A sample site index card is shown below in Table 2.

Table 2- Site Index Card Records

	Site Information:
1.	URL for the Site directory for which this information applies
2.	Top URL for this Site
3.	INCLUDE or EXCLUDE URL
4.	Language
5.	Site Name _____ (RealName)
6.	Site Description _____ (limited to 25 words)
7.	Name of the site Host _____
8.	Contact Name
9.	Address
10.	Phone
11.	Fax
12.	Technical Contact name:
13.	Technical Contacts' direct phone number
14.	Technical Contacts' email address
15.	Would you like the Technical Contact to receive email notification of every successful site index update.
16.	Business Contact name:
17.	Business Contacts' Direct phone number
18.	Business Contact's email address
19.	Type of site
20.	Company - Name _____
21.	Organization Name _____
22.	Individual - Name _____
23.	Category
•	• General
•	• Specific Category
•	• Special interest
24.	Related categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
25.	Demographics Site's intended audience

•	• Age
•	• Sex
•	• etc.
26.	Location of Site's intended audience:
•	• World
•	• Country
•	• State or Province
•	• City
•	• District
27.	Key words (repeated words will not be indexed)
28.	Key Phrases (repeated phrases will not be indexed)
29.	Related Site's
30.	Comments
31.	Type of products for sale _____
32.	Location of products database _____
33.	Type of database _____ SQL or ? or ?
34.	Rating
35.	Rating Descriptors
36.	Others to be added.

The web site owner can create a different site index card for each page or directory on the site. For example, if the web site includes pages in different languages, the web site owner should create a site index card for each language with keywords and categories that match the language. Once the web site owner has filled in the index card form, they will click a button on a web page from the web server at the central server, and a web server creates an encoded html file that is then sent or downloaded to the site owner's computer. Each encoded index card file could be given a particular name, such as index card-domainname-com-directory-directory-directory.html, and the site owner is instructed to copy the encoded file into the specified web directory on the site.

At anytime, the web site owner can visit the central server site, update their index card, and download a new encoded index card. When updating an existing index card, the current index card information for the URL entered is displayed to reduce input time. Any site index card will supercede the host index card information, and information contained in the site index card will be assumed to be more current and accurate and will be used by the agent for indexing purposes. A site index card that is farther down in the directory tree from the root

directory will supercede a site index card that is above it in the directory tree. A site owner can request that their web site be excluded from the index by checking the EXCLUDE box next to the URL and copying the index cards into the directory to be excluded.

If the host or web site URL is not currently being indexed, the web server
5 performs the following operations. First, an automatic email is sent to contacts at the host to encourage the host to install the agent. An automatic email is also sent to a contact person for the web site with a "Thank You" and a request that they ask their host to install the agent. In addition, a retrieval order is generated for the central server to retrieve the index card file from the web site in one hour. If the retrieval order is unsuccessful, it will be repeated 2, 4, 8, 24 and
10 48 hours later, until successful. If still unsuccessful after 48 hours, the retrieval order is canceled. By verifying the presence of the site index card in the specified location, unauthorized information about a site may not be created by a third party in an attempt to have their site indexed along with a more popular site. This is a common problem with existing search engines where a third party copies the keywords from a meta tag in a popular site. The bogus site with
15 copied keywords is then submitted to a search engine for indexing, and when search queries are applied to the search engine that produce the popular site the bogus site is also produced. This may not be done with the site index card because the index card is not an html page available to outside persons and because it is encrypted so even if the file is obtained the information contained therein is not accessible.

Index Card Verification and Indexing by the Agent

Software to create index cards and agent programs is distributed free to software
25 publishers for inclusion in their web authoring software and to web server manufactures, publishers and OEMs for pre-loading on or inclusion with their products.

As mentioned above, the agent 204 is a software program that a web host downloads from the web servers 214 and installs on the host's server. To install the agent 204, the host runs an agent installation program, which collects information about the web site host

and about the site itself, and also creates the web site host's index card 206 of non-HTML information. As part of the installation, the site host schedules a preferred time of day for the agent 204 to automatically index the web site and transfer index updates to the central server 202.

- 5 Although the agent 204 automatically indexes the sites on the host at scheduled times, the host can, at anytime, initiate an indexing update by opening the agent 204 and manually initiating an index update.

 The agent 204 updates the index of the local site. First, the agent 204 determines whether the last index update was completed and transmitted successfully. If not, the agent 204
10 renames the Old—Site-Index file to Site-Index and the Old-Site-File-List to Site-File-List. The agent 204 then calculates a digital signature for the Site-Index file and a signature for the Site-File-List file and compares each to the digital signatures created at the end of the last successful update for Site-Index and Site-File-List files. If the digital signatures match, the agent 204 sends them to the central server 202 for comparison and waits for confirmation.

- 15 If the central server 202 does not confirm the match of the digital signatures (i.e., the signatures for the Site-Index and Site-File-List files on the central server 202 do not match those on the remote site), the agent 204 deletes the Site-Index and Site-File-List files, and notifies the central server 202 to delete all site records. Next, if the agent 204 was updated and Fields were added or deleted from the Site Index file, then the agent updates the Site-Index file to
20 include the updates. The agent 204 then determines if the Site-File-Lists file exists, and renames the Site-File-List file to Old-File-List and create a text file named Site-File-List. If no Site-File-List exists but Old-File list exists, the agent 204 copies the Old-File-List file to Site-File List. If no Site-File-List and no Old-File-List files exist, the agent 204 creates a text file named Site-File-List. The agent 204 then calculates a digital signature hash for each file on the site and the
25 host index card and records the file name including full path and digital signature hash of all files.

 If the central server 202 verifies that the digital signature hash of the Site-Index file and the digital signature hash for the Site-File-List file match, the agent 204 verifies the index card files. More specifically, the agent 204 determines if the file index-card.html file name

does not match the directory in which it is located. If the file index-card.html is not in the expected directory, the agent 204 sends a warning email to the site contact listed in the index card and then renames index card.html to WrongDirectoryindex-card.html.

If the agent 204 determines that all index-card.html files match the directory in which they are located, the agent 204 deletes a file named Exclude-File-List, creates a text file named Exclude-File-List, checks index cards for EXCLUDE sites flags, and adds file names of files to be excluded from the index to the Exclude-File-List file. The agent 204 then creates a Deleted-File-List file containing a list of files that no longer exist on the site in their original location. More specifically the agent 204 deletes the old Deleted-File-List file, creates a text file called Deleted-File-List, compares the Site-File-List file to Old-File-List file and records in the Deleted-File-List any files in the Old-File-List that are not in Site-File-List.

The agent 204 then creates a New-File-List file containing a list of files that were created or modified since the last update. Next, the agent 204 indexes the corresponding site and creates a new Site-Index file.

The agent 204 then creates a New-Records-Index file and a Deleted-Records-List file. The agent 204 next removes records of deleted or modified files from the Site index.

The agent 204 then runs an indexing program against all files in the New-File-List file and creates a record for each new key word, phrase, MP3, Video, Movie, Link and index card information and adds these to the Site-Index file. The agent 204 then copies each new record created to the New-Records-Index file. If new fields were added to the Site Index, the agent 204 runs the indexing program against all files for new field information and creates records in Field-Update-Index for all information found. The agent 204 then updates the Site-Index file from the Field-Update-Index file.

At this point, the Site-Index file has been updated, and the agent 204 calculates a digital signature for the Site-Index file. Finally, the agent 204 creates a Site-Map file for the sites serviced by the agent. Once the Site-Map file has been generated, the agent 204 sends New-Records-Index and Deleted-Records-List files to the central server 202.

The agent 204 generates a site index, which is a database. The database includes a number of tables, each table consisting of records (rows) and fields (columns). Each table in the

database includes similar records to speed searches. All Tables may be sorted alphabetically and then by category. In one embodiment of the agent 204, the agent generates Tables 3-12 as shown below.

5

Table 3-Agent Created Keywords Table Fields

1.	i. Keyword
2.	Category –General, Specific, Special Interest Categories
3.	Related categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
4.	Host IP address,
5.	Site URL,
6.	Unique Record Identifier
7.	Location of first occurrence of word
8.	URL for first occurrence of word
9.	Number of occurrences of word
10.	Does word appear in meta header
11.	Does word appear in index card keywords
12.	Does word appear in index card phrases
13.	Demographics restrictions (Y or N)
14.	Location restrictions (Y or N)
15.	Date file containing Key Phrase was created
16.	Link to Site index card
17.	Link to Host index card
18.	Link URL Link Table
19.	Html tag information
20.	XML tag information
21.	Ranking

Table 4-Agent Created Key Phrases Table Fields

1.	ii. Key Phrase
2.	Category – three letters representing General, Specific Special Interest, and Categories
3.	Related Categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
4.	Host IP address,
5.	Site URL,
6.	Unique Record Identifier
7.	Location of first occurrence of Phrase
8.	URL for first occurrence of Phrase
9.	Number of occurrences of Phrase
10.	Does Phrase appear in meta header
11.	Does Phrase appear in index card phrases

12.	Demographics restrictions (Y or N)
13.	Location restrictions (Y or N)
14.	Date file containing Key Phrase was created
15.	Link to Site index card
16.	Link to Host index card
17.	Link URL Link Table
18.	Html tag information
19.	XML tag information
20.	Ranking

Table 5-Agent Created Products Catalog

1.	iii. Type of product
2.	Category – three letters representing General, Specific, and Special Interest Categories
3.	Related Categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
4.	Product description
5.	Site URL,
6.	Unique Record Identifier
7.	iv. Product Number
8.	v. Product price
9.	vi. Feature or option
10.	Feature or option
11.	Feature or option
12.	Link URL Link Table

Table 6-Agent Created Articles & Documents Table

1.	vii. Type of Articles or Documents
2.	Category – three letters representing General, Specific, and Special Interest Categories
3.	Related Categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
4.	Subject of Articles or Documents
5.	Site URL,
6.	Unique Record Identifier
7.	viii. Date
8.	ix. Author
9.	x. Source of Articles or Documents
10.	
11.	

12.	Link URL Link Table
-----	---------------------

Table 7-Agent Created MP3 Table

1.	xi. Title of Song
2.	Category – three letters representing General, Specific, and Special Interest Categories
3.	Related Categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
4.	Host IP address,
5.	Site URL,
6.	Unique Record Identifier
7.	xii. Name of Group
8.	xiii. Name of Artist
9.	xiv. Name of Artist
10.	Name of Artist
11.	Name of Album
12.	Name of Record label
13.	Name of producer
14.	Name of MP3 file
15.	Size of MP3 file
16.	Year produced
17.	Link to Site index card
18.	Link to Host index card
19.	Link URL Link Table

Table 8-Agent Created Video Table

1.	xv. Name of Video
2.	Category – three letters representing General, Specific Special Interest Categories
3.	Related Categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
4.	Host IP address,
5.	Site URL,
6.	Unique Record Identifier
7.	Artists name 1
8.	Artists name 2
9.	Artists name 3
10.	Name of director
11.	Year produced
12.	Name of Studio
13.	Name of producer
14.	Size of file
15.	Link to Site index card
16.	Link to Host index card

17.	Link URL Link Table
-----	---------------------

Table 9-Agent Created URL Link Table

1.	xvi. URL link
2.	Category – three letters representing General, Specific Special Interest Categories
3.	Related Categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
4.	Host IP address,
5.	Site URL,
6.	Unique Record Identifier
7.	URL link to other links in the Link Table.
8.	Other desired information
9.	
10.	
11.	

Table 10-Agent Created Site Index Card Table

1.	Site URL
2.	Site Name _____ (RealName)
3.	Site Description (limited to 25 words)
4.	Name of site Host _____
5.	Contact Name
6.	Address
7.	Phone
8.	Fax
9.	Contact person for technical related issues:
10.	Contacts' direct phone number
11.	Contacts' email address
12.	Contact person for business related issues:
13.	Contacts' Direct phone number
14.	Contact's email address
15.	Type of site
16.	Company – Name _____
17.	Organization Name _____
18.	Individual – Name _____
19.	Category
20.	General
21.	Specific Category
22.	Special interest
23.	Related Categories 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
24.	Demographics Site's intended audience
25.	Age
26.	Sex

27.	Location of Site's intended audience:
28.	World
29.	Country
30.	State or Province
31.	City
32.	District
33.	Key words (repeated words will not be indexed)
34.	Key Phrases (repeated phrases will not be indexed)
35.	Related Site's
36.	Comments
37.	Others to be added.

Table 11-Agent Created Company Information from Host Index Card

1.	IP number
2.	Domain Name Server
3.	Type of Domain Name Server
4.	ISP - Name _____
5.	Company - Name _____
6.	Individual - Name _____
7.	ISP name
8.	Company Name if different
9.	Contact Name
10.	Address
11.	Phone
12.	Fax
13.	Contact person for technical related issues:
14.	Contacts' direct phone number
15.	Contacts' email address
16.	Contact person for business related issues:
17.	Contacts' Direct phone number
18.	Contact's email address
19.	General Information: (optional)
20.	Area served
21.	Number of email boxes hosted _____ or N/A
22.	Number of Domain Names hosted(?) _____ or N/A
23.	Number of web sites hosted. _____ or N/A
24.	Other Desired Information

Table 12-Agent Created Site Map

1.	Site Map
2.	IP number

Updating the Central Copy of each Index Card

Figure 3 is a bubble chart illustrating the process executed by the queue manager 302 of Figures 1 and 2 in queuing update entries and transferring these entries to the remote queue manager 304. The queue manager 302 receives update entries 600 from the agent update server 222 and update entries 602 from the index card server 228, and places these update entries in an update queue 604. The entries in the queue 604 are transferred to a queue database 606. Once the queue 604 is done receiving update entries 600, 602, which may be when the queue is full or at predetermined intervals, the queue manager 302 goes to step 608 and retrieves the queue entries from the queue database 606 and sends them to the remote queue manager 304. The update entries stored in the queue database 606 are thereafter processed by the update processing server 306 (see Figure 1) to update the local master index on master index sever 218 (see Figure 1). The queue manager 302 also receives a deletion request (not shown) from the update processing server 306 and deletes update entries stored in queue database 606 in response to this deletion request.

Figure 4 is a bubble chart illustrating the overall data flow between the search engine 202, agent, and index card components of the active indexing system. Each aspect of the overall data flow has already been described in a corresponding section above, and thus Figure 4 will now be described merely to provide a brief description of the data flow of the indexing system according to one embodiment of the present invention.

The queue manager 302 receives updates from each of the agents 204 and from the index card check server 228, which services sites without an agent 204 as previously described. The queue manager makes update and deletions to the queue database 602 corresponding to the received updates, and also provides a mirror copy of these updates to the remote queue manager 304. The update processing server 306 retrieves the update entries from the queue manager 302, and applies the updates to the master index servers 218. The server 218 updates the master index to include the applied updates, and the update processing server 306 then sends a deletion request to the queue manager 302 to delete the corresponding entries from the queue database 602.

Once the master index server 218 has updated the master index, the server updates the segmented index stored on the search index servers 216, 217 as previously described. Each of the search index servers 216, 217 updates its corresponding portion of the segmented index in response to the updates from the master index server 218. As previously mentioned, the entire segmented index stored on the index servers 216 is continuously available for processing search requests even during updating of the segmented index. The entire segmented index is available due to the redundant architecture of the servers 216, 217, as previously described.

Use of Index Cards to Limit Exposure of Objects

With the architecture of the indexing system, everything, including each field in a local database, is treated as an object. Instead of copying each object to a central location, an object reference is created at each local site and sent to a cataloging location or locations. The objects are not duplicated in a monolithic central database. One advantage to this architecture is that the decision of whether to expose the existence and classification of local objects becomes the responsibility and choice of the author, rather than a generic decision. In the system, the implementation of retention rules and the physical location of the objects remain with the author. The searchable central catalog merely references the distributed objects, eliminating the need to make full copies and therefore manage a large storage system.

Rating with Agents or Index Cards

The agent 204 may also generate ratings for objects stored on the associated sites so that users may filter their searches based upon the generated ratings. For example, in one embodiment, an owner of a web site provides a rating of his site, such as a "G," "R," or "X" rating. In addition, the web host, on which the agent 204 runs, also provides a rating that the host believes applies to the site. The agent 204 then parses the pages on the site and looks for adult content "trigger" words, such as "XXX" or "XXX-Rated." If the agent 204 finds enough occurrences of such trigger words, the agent "flags" the web site for review to determine the correct rating for the site. To rate the site, the agent 204 compares the words in the web pages to words in a list of ratings values.

The list of ratings values may be, for example, words that are generally found on adult web sites, such as profane and sexually explicit words. The list of ratings values may be

generated by a human or may be automatically generated by the agent 204. To automatically generate the list, the agent 204 could, for example, parse known adult web sites. Such known adult web sites could be identified by determining those sites in the catalog that include the phrases "adult content" or "X-rated." Once these sites are identified, the agent parses the pages and determines frequently used words on such pages, and may also determine the frequency with which such words occur on these pages. The frequently used words and associated frequencies are then compiled to form the list of ratings values.

After flagging web sites for review, the review may be either through human review of the web site or through automated review performed by the agent 204. In automated review of flagged web sites, the agent 204 could, for example, determine the frequency of occurrence of words in the list of ratings values, and then set the rating of the web site as a function of the frequency. For example, if the frequency is greater than some threshold T1, the web site is rated "R," and if greater than a second threshold T2, where $T2 > T1$, the site is rated "X."

One proposed system for rating web pages on the Internet is described in *A Best Practices Model* by Members of the Information Society Project at Yale Law School, J.M. Balkin, Beth Simone Noveck, Kermit Roosevelt (July 15, 1999), which may be found at <http://webserver.law.yale.edu/infosociety/>. In this proposed system, three layers are implemented to provide for rating web pages. The first layer includes a basic vocabulary of, for example, thirty to sixty terms that are used in rating a web page by a first party, typically the site owner containing the web page. The second layer includes rating templates developed to reflect a particular ideology. Third parties, such as the NAACP or Christian Coalition, would develop such templates to reflect a particular value system. The templates would include terms in the basic vocabulary being categorized and scalar values assigned to each item to reflect the value system. Finally, in layer three, individuals could customize or modify a template to suit their individual values. For example, a template developed by the Christian Coalition could be further modified to include scalar values for web sites designated as racist by the NAACP.

The indexing system can utilize such a rating system to perform filtering of search results at the central server 202. For example, user's browsers can be registered with the central

server 202, and part of this registration includes selection of a template and any desired modifications to the selected template. Thereafter, whenever the user's browser applies a search query to the central server 202, the browser registration is identified and the search results generated in response to the query are "filtered" according to the template and any template modifications associated with the registered browser.

The indexing system also may perform adult-content locking. In conventional search engines, adult-content web sites are automatically provided in response to applied search queries. The only way for a user to filter adult-content is through a filter on his browser. Thus, current search engines are "opt-in" in that the search engine does not preclude adult-content pages from being returned in response to applied search queries. Conversely, in one embodiment of the indexing system, the user is automatically opted out of receiving adult-content web pages in response to applied search queries. The user must reverse this default "opt-out" status and elect to receive adult-content web pages in the system. This could be done, for example, by registering a browser with the system so that, when the registered browser is identified, adult-content web sites will be returned in response to applied search queries. Alternatively, a machine level lock using the computer or machine identification, such as the CPU or Windows identification number, could be utilized. In this approach, regardless of the browser being utilized on the computer, adult-content is either returned or not returned in response to applied search queries. This approach may be particularly desirable for parents who want to preclude their children from accessing adult-content, since a child cannot merely use a new browser on the same machine and thereby circumvent the filter the parent has on his or her browser.

Ranking with Index Cards or Agents

The indexing system may also perform ranking of web pages having references in the central index. First, the agent 204 may perform positional and contextual rankings for particular words in the web pages on a site. The positional rankings assign a ranking value to a word based upon, for example, the location of the word in the web page and the position of the word relative to other words in the page. The contextual ranking is determined using contextual information about the site contained in the index card 206. For example, if a word in a web page

corresponds to a category as listed in the index card 206 (see Table 2) , the word will be assigned a higher ranking.

In addition to rankings generated by the agent 204, the central server 202 also generates rankings for the central index. For example, the central server 202 may generate rankings based upon whether a page is a source or reference to the desired data. In such an analysis, the number of links inbound to the page being processed (and the content of these external pages) are used to generate a ranking value for the current page. Similarly, an analysis of the links outbound from the page are performed, determining the importance of the page as a "reference" to information as opposed to being a "source" containing information referenced by other pages. As a crude example, the number of inbound links to a page would determine it's ranking. Rankings may also be determined based upon user input such as the usage or popularity of a site as measured by how often the site is linked as the source site in other sites, or through positive comments entered by users about the context or ranking of a site.

All the methods of ranking just described are known as static rankings, meaning that the ranking is determined before a particular search query is applied. In addition to static rankings at the central server 202, the central server may also perform dynamic ranking of search results. Dynamic rankings are a function of the applied search query, are not predetermined, and are not independent of the query. For example, if the applied search query is "red barn," the word "barn" is probably more important than "red" so search results including the word "barn" will have their ranking increased relative to those containing only the word "red."

Furthermore, ratings could be applied to search queries to create another type of dynamic ranking at the central server 202. For example, a rating could be used as part of a query, to restrict searches to those sites or documents which have a certain minimum or maximum rating. Once the result set of matching sites or documents is retrieved, the resulting list could be ranked in order of rating value. For example, one might search for all pages with a "PG" rating (as determined by the Motion Picture Association Of America) or lower (where "G" would be the lowest). Pages could then be ranked where "G" rated pages are ranked higher than pages which have more mature content.

Finally, a user may select which ones of the previous methods of rankings should be applied in ranking search results generated in response to his applied query. For example, a user could specify that his search results are to be ranked only on the basis of popularity, or only on the basis of positional and contextual rankings and the applied search query. Another
5 example for the use of dynamic ranking is, using the information in the index card 206, the search results can be ranked dynamically based on the geographic distance from the searcher.

Index cards may be read by external systems as opposed to local programs such as an agent. The information retrieved from the index cards may then be used to perform ranking analysis on the pages related to that index card. For example, the physical location data (address
10 or geographical coordinates) might be used by a system to rank sites based on the proximity of the site owner to a major population center. In the current industry environment, there are automated search and collating systems which regularly interrogate or visit websites and extract relevant information from those sites. The resulting information is then analyzed and presented in some form to subscribers of a service or members of a "portal" site. Similar methods could be
15 used to regularly retrieve index cards from known web sites and then perform analysis on the information contained in those index cards.

Ranking with Relationship Values Calculated by Agents

According to another embodiment of the agent 204, the agent calculates a value representing the distance in text between objects and thereby determines which objects at a site
20 are most likely to relate to each other. At the catalog site, these relationship values are combined with the relationship values from other sites to create a relationship value table. This relationship value table represents the likelihood of an object occurring together with another object. This table may be used to refine searches and create relevance ranking.

Clustering Search Results with Index Cards

25 The index cards allow documents on a web site to be automatically categorized based on classifications entered by the web site administrator. Such classifications can apply to the entire site, certain subsets of the site content, or individual documents stored on the site. This categorization information is transmitted by the agent software to the central index or retrieved by the central server in an embodiment without agents.

Since the classification information is known for all pages, the results of a search performed at the central index can be clustered by category. It is not even necessary to display the list of matching documents -- the user of the system can simply view the number of documents found within each category. The results display can further be refined by structuring the categories in a hierarchical fashion, where selection of a category containing matches would cause a display of all subcategories containing matches as well as any matches found within the selected category itself.

The primary advantage of such clustering is to provide the user of the system context upon which to base a decision as to the relevance of the set of documents found. For example, a query for "Ford" might produce matching documents within the topics of "Retailing", "News" and "Reference Material." If the searcher was interested in finding a Ford Motor Company retail dealership, matches under "Retailing" would be of more interest than all others. By providing the context indicating in what category the search results fall, the searcher is able to discern what search results are relevant without the use of ranking algorithms.

Example part I:

The search for "Ford" found the following:

Retailing:	700 matches in 15 subcategories
News:	3000 matches in 8 subcategories
Reference Material:	36 matches in 2 subcategories

Example part II

Retailing (700 matches)

Ford Motor Company Retail Dealers:	550 matches in 51 subcategories
Repair Facilities – Ford Motor Co. Products:	150 matches in 12 subcategories

In the example part I above, a search has been performed by the user for the word "Ford." A resulting display containing information about matching pages has been generated and is now displayed on the user's screen within a standard web browser. Each heading listed is a hyperlink which, when selected, causes the list of matching pages and subcategories to be displayed as is shown in the example part II where the user has selected the hyperlink titled "Retailing" in part I. In both parts, the categorizing of results based on data supplied by index cards is shown.

This example shows that categorization can be done at more than one level in a hierarchy or ontology. Thus, a page which was categorized as “Retailing / Automobiles / Ford Motor Company / Dealers / USA / California / Los Angeles Area” would actually apply towards the match count at several levels of a hierarchy.

5 Implementation with Java

A Java-based implementation of the software elements of the indexing system is now described according to one embodiment of the present invention.

First, most users, with the exception of individuals using the search engine to find information on the Internet, will need to be authenticated, as illustrated in Figure 12. The operations that need to be supported to authenticate users include those shown in Table 13.

Table 13

Case	Description
Register	This requires the entry of sufficient information for central records. Index card users can register themselves and have restricted access only to their own records. .
Logon	Allows the user to access the part of their system they have permission to access.
Logoff	Ensures the user is no longer active on this system.

A web site owner may be a content developer that owns a web site and develops content for that web site. Such a person may be hosting their own site but is often simply deploying content on another entity's machine. In this role, the owner does not administer the hardware/software and manages only site content. The operations that need to be supported include those shown below in Table 14.

Table 14

Case	Description
List Existing Index cards	Provides a selection list of existing index cards previously created by this user.
Create Index card	Create a new index card, name it and fill out a form to describe site content.
Edit Index card	Created index card can be edited to keep information current or to correct errors.
Duplicate Index card	An existing index card can be used as a starting point for a new index card. It must be given a new name when duplication occurs.
Delete Index	An outdated index card can be removed from the database.

card	
Download Index card	A index card must be downloaded to the user's system in the format required for deployment.
Install Index card	A index card must be installed on the web site after it has been downloaded.
Configure Notification	Notification is sent to the user when the agent updates the index or when index cards are scanned by the update manager. The user can choose the level of notification they receive.

The role of the web searcher is the end user of the system, which operates for free on the Internet, or any other future global communications network, since revenue is brought in through other entities. A web searcher expects to use the system to find things and navigate through their findings. Operations that need to be supported for web searches include those shown in Table 15.

Table 15

Case	Description
Select Categories	High-level categories are presented for user selection.
Execute Search	This means entering one or more keywords and executing the actual search.
Refine Search	After a preliminary search, additional options are made available to narrow the search to more targeted information.
Select Ranking Algorithm	Once the number of hits in the result set has been reported, the user can choose the algorithm to rank results by.
Display Results	After selecting a ranking algorithm, the user can display results. After seeing the first result page, the user can choose to continue to the next page or go back at will.
Request Site Reclassification	If one of the resulting page hits from a search appears questionable to a user, they may sent a request that we reclassify the site. This is especially important with adult content mistakenly found in non-adult searches.

Persistent information is stored in a database or in temporary files on the system.

- 10 The database tables relate to each other as shown in Figure 5. The following tables specify the data dictionary for tables illustrated in Figure 5 that relate to index cards, ranking, and rating.

The ContextTable 2111 of Figure 5 defines possible contexts for indexing. These represent a short list such as HEAD, BODY and META for HTML pages. The ContextTable does not reference any other tables.

The PopularityTable 2105 of Figure 5 stores statistical information about the popularity of given pages, based on how often the page was selected after a search. This information is used when the user chooses popularity-based ranking at result display time. The popularity table is shown in more detail in Table 16 below. The PopularityTable 2105 references the PageTable 2103.

Table 16

Field	Type	Key	Description
PopularityID	LONG	Primary	This is the primary key for the PopularityTable. It is an auto-increment integer value that is unique across the system.
PageID	LONG	Foreign	This is a foreign key pointing to the PageTable's primary key.
ClickCount	LONG	None	The number of times we have recorded users clicking through this page on the search engine site.

The IndexCardTable 2121 of Figure 5 stores information related to electronic index cards. The IndexCardTable references the FieldTable 2117, the ContactTable 2123, the GeographyTable 2119, the DemographicTable2125, the RatingTable 2129, and the DatabaseTable 2127. The fields of the IndexCardTable are shown in Table 17 below:

Table 17

Field	Type	Key	Description
IndexCardID	LONG	Primary	This is the primary key for the IndexCardTable. It is an auto-increment integer value that is unique across the system.
SiteName	TEXT (64)	None	Name of site to be indexed.
RootURL	VARCHAR	None	URL for the root of the site structure to be indexed with this index card.
ExcludeURLList	VARCHAR	None	List of URL within this structure to be excluded.
OwnerContactID	LONG	Foreign	This is a foreign key pointing to the ContactTable's primary key.
BusinessContactID	LONG	Foreign	This is a foreign key pointing to the ContactTable's primary key.
TechnicalContactID	LONG	Foreign	This is a foreign key pointing to the ContactTable's primary key.
DBAContactID	LONG	Foreign	This is a foreign key pointing to the ContactTable's primary key.
CategoriesID	LONG	Foreign	This is a foreign key pointing to the

			CategoriesTable's primary key.
RatingID	LONG	Foreign	This is a foreign key pointing to the RatingTable's primary key.
DemographicID	LONG	Foreign	This is a foreign key pointing to the DemographicTable's primary key.
GeographyID	LONG	Foreign	This is a foreign key pointing to the GeographyTable's primary key.
CustomFieldList	VARCHAR	None	List of user-defined fields to be indexed by the agent.
DatabaseID	LONG	Foreign	This is a foreign key pointing to the DatabaseTable's primary key,

The GeographyTable 2119 of Figure 5 is shown in more detail in Table 18 below. This table stores information related to the geographical target market for a given index card. The GeographyTable does not reference any other tables.

Table 18

Field	Type	Key	Description
GeographyID	LONG	Primary	This is the primary key for the GeographyTable. It is an auto-increment integer value that is unique across the system.
TargetCountry	TEXT (64)	None	Target market country.
TargetState	TEXT (64)	None	Target market state.
TargetCounty	TEXT (64)	None	Target market county.
TargetCity	TEXT (64)	None	Target market city.
TargetCommunity	TEXT (64)	None	Target market community name.
TargetStreet	TEXT (64)	None	Target market street.

The DemographicTable 2125 of Figure 5 is shown in more detail in Table 19 below. This table stores information related to the demographic target market for a given index card. The DemographicTable does not reference any other tables.

Table 19

Field	Type	Key	Description
DemographicID	LONG	Primary	This is the primary key for the DemographicTable. It is an auto-increment integer value that is unique across the system.
AgeFrom	INT	None	Demographics minimum age.
AgeTo	INT	None	Demographics maximum age.
RevenueFrom	MONEY	None	Demographics minimum revenue.
RevenueTo	MONEY	None	Demographics maximum revenue.

The RatingTable 2129 of Figure 5 is shown in more detail in Table 20 below. The table stores information related to ratings associated with a given index card. The RatingTable does not reference any other tables.

Table 20

Field	Type	Key	Description
RatingID	LONG	Primary	This is the primary key for the RatingTable. It is an auto-increment integer value that is unique across the system.
RatingName	TEXT (64)	None	Name of this rating system.
DimensionCount	INT		Number of dimensions in the rating system.
ValueListList	VARCHAR		List of values for each dimension.

The packages in the system are either low-level utility packages, components, applications and services or user interface elements. Support packages, like the database, catalog, html and xml packages, provide a foundation for other program functionality. A few of the services, the message and index services, for example, are grouped as shared because several of their classes provide functional capabilities between both the agent and server elements. The index card package is also shared. The application and service level packages construct the agent and the server-side Jini services. Taken together, the classes in these packages function together as a complete, integrated, distributable system.

The shared index card package is shown in more detail in the following

Table 21

Class	Description
IndexCardService	This is an implementation of a service that provides access to index cards on the server. It is used by the servlets to provide index card management services and by the update manager to verify content.
Index cardDocument	An index card document is an XML representation of an index card.
DocumentIndexCard	A document index card applies to html documents.
DatabaseIndexCard	A database index card applies to databases on the target machine.

The shared rating package contains classes related to rating systems, as shown in the following Table 22.

Table 22

Class	Description
Rating	Interface defining methods required to operate as a rating class.
AbstractRating	Abstract class which implements common behavior and internal methods useful in implementing rating classes.
MovieRating	Movie rating class offers two option dimensions (G, PG, R, X... and Sex, Violence...). The set of options is retrieved using the indexed getOptionList method.
InternetRating	This class is a place-holder for future rating systems.

Each of the previously recited methods is performed by a program contained on a computer-readable medium, such as a CDROM. The program may also be contained in a computer-readable data transmission medium that may be transferred over a network, such as the Internet. The data transmission medium may, for example, be a carrier signal that has been modulated to contain information corresponding to the program.

It is to be understood that even though various embodiments and advantages of the present invention have been set forth in the foregoing description, the above disclosure is illustrative only, and changes may be made in detail, and yet remain within the broad principles of the invention. Therefore, the present invention is to be limited only by the appended claims.

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

We claim:

1. A method of making categorization information about a site on a computer network available to other computer systems on the network, comprising:

- a) creating a site on a computer network comprising a plurality of objects which bare
5 available for access by other computer systems on the network;
- b) creating a data file associated with said site containing human provided categorization information about said site; and
- c) making the contents of said data file available to other computer systems on the network.

10 2. A computer data set containing a computer program which, when run on one or more computer systems, causes the systems to perform the method of claim 1.

3. The method of claim 1 further comprising assembling contents from said data file and a plurality of similar data files for other sites and performing searches of said assembled contents to find desired sites.

15 4. The computer data set of claim 2 further comprising the limitations of claim 3.

5. The method of claim 3 further comprising structuring categories for said categorization information in a hierarchical fashion with subcategories, where selection of a category for a search causes results of said search to include sites associated with subcategories of said category.

20 6. The computer data set of claim 2 further comprising the limitations of claim 5.

7. The method of claim 1 where said categorization information comprises a human language in which objects of the site are presented when accessed by other computer systems on the network.

8. The computer data set of claim 2 further comprising the limitations of claim 7.

9. The method of claim 1 where said categorization information comprises a subject matter to which objects of the site relate as perceived by humans when accessed by other computer systems on the network.

10. The computer data set of claim 2 further comprising the limitations of claim 9.

5 11. The method of claim 1 where said categorization information comprises a geographic location to which objects of the site relate as perceived by humans when accessed by other computer systems on the network.

12. The computer data set of claim 2 further comprising the limitations of claim 11.

10 13. The method of claim 1 where said categorization information comprises a general description of a target audience by whom an owner of the site wishes the site to be accessed.

14. The computer data set of claim 2 further comprising the limitations of claim 13.

15 15. A method of constructing a catalog of rankings from objects stored within a network, the network including a plurality of interconnected computers with one computer storing the catalog and being designated a cataloging site and each of the other computers storing a plurality of objects and being designated a source site, the method comprising:

a) running on the cataloging site a program which assembles data relating to objects stored on the two or more source sites where, for each of the two source sites, such data is gathered from a file that is not a part of any of said objects and said file contains data entered by a human about at least one of said objects; and

20 b) ranking at least some of the assembled data as a function of a set of ranking rules, thereby assigning rankings to the assembled data to generate the catalog of rankings where at least one of the rankings has a value that is function of said human input data about one or more objects with which the ranking is associated.

25 16. A computer data set containing a computer program which, when run on one or more computer systems, causes the systems to perform the method of claim 15.

17. The method of claim 15 wherein the data assembled at the two or more source sites is assembled by an agent program running on the source site.

18. The computer data set of claim 16 further comprising the limitations of claim 17.

19. The method of claim 15 wherein at least one of the rankings has a value that is
5 function of human usage of the object references.

20. The computer data set of claim 16 further comprising the limitations of claim 19.

21. The method of claim 15 wherein some of the assembled data comprises data from the contents of objects on at least one of the source sites.

22. The computer data set of claim 16 further comprising the limitations of claim 21.

10 23. The method of claim 15 wherein some of the assembled data comprises meta data relating to objects on at least one of the source sites.

24. The computer data set of claim 16 further comprising the limitations of claim 23.

25. The method of claim 15 wherein some of the assembled data comprises ratings of objects on the source site.

15 26. The computer data set of claim 16 further comprising the limitations of claim 25.

27. A method of constructing a catalog of rankings from objects stored within a network, the network including a plurality of interconnected computers with one computer storing the catalog and being designated a cataloging site and at least two other computers storing a plurality of objects and being designated source sites, the method comprising:

20 a) running on each source site a program which assembles data relating to objects stored on the source site;

b) ranking at least some of the assembled data as a function of a set of ranking rules, thereby assigning rankings to the assembled data;

c) transmitting the rankings from each source site to the cataloging site; and

d) aggregating the rankings at the cataloging site to generate a catalog of rankings.

28. A computer data set containing a computer program which, when run on one or more computer systems, causes the systems to perform the method of claim 27.

5 29. The method of claim 27 wherein at least one of the rankings has a value that is function of human input data about one or more objects with which the ranking is associated where the human input data is stored in a file on the source site which file is not a part of said one or more objects and assembled by said program.

30. The computer data set of claim 28 further comprising the limitations of claim 29.

10 31. The method of claim 27 wherein some of the assembled data comprises data from the content of objects on the source site.

32. The computer data set of claim 28 further comprising the limitations of claim 31.

33. The method of claim 27 wherein some of the assembled data comprises meta data relating to objects on the source site.

15 34. The computer data set of claim 28 further comprising the limitations of claim 33.

35. The method of claim 27 wherein each transmitted ranking is accompanied by a command to the cataloging site instructing the cataloging site what to do with the ranking.

36. The computer data set of claim 28 further comprising the limitations of claim 35.

20 37. The method of claim 27 wherein the program further assembles object references for objects on the source site, and these object references are transmitted to the cataloging site and aggregated into the catalog on the cataloging site.

38. The computer data set of claim 28 further comprising the limitations of claim 37.

39. The method of claim 27 wherein the program further transmits to the cataloging site some of the assembled data which is aggregated into the catalog on the cataloging site.

25 40. The computer data set of claim 28 further comprising the limitations of claim 39.

41. The method of claim 27 wherein at least one of the rankings relates to a set of objects on the source site.

42. The computer data set of claim 28 further comprising the limitations of claim 41.

43. The method of claim 41 wherein the agent calculates a relationship value representing a distance in text between objects and, at the cataloging site, these relationship values are combined with relationship values from other sites to create a relationship value table representing the likelihood of an object being similar to another object.

44. The computer data set of claim 42 further comprising the limitations of claim 43.

45. A method of rating objects stored at a site on a network and constructing a catalog of ratings, the network including a plurality of interconnected source computers and a central computer with access to the objects, the method comprising:

a) running on a central computer a program which processes objects stored on the source computers, thereby assembling values found in at least one of the objects for comparison to a list of rating values;

b) generating a rating for each object by relating the values found in the object to a list of human input rating values supplied by an owner of the site and stored in a file associated with the site which file is read by said program; and

c) aggregating the ratings to generate the catalog of ratings.

46. A computer data set containing a computer program which, when run on one or more computer systems, causes the systems to perform the method of claim 45.

47. The method of claim 45 wherein each of the rating values comprises a word.

48. The computer data set of claim 46 further comprising the limitations of claim 47.

49. The method of claim 45 wherein additional human input rating values are supplied by a host of the site and stored in said file.

50. The computer data set of claim 46 further comprising the limitations of claim 49.

51. A method of rating objects stored at a site on a network and constructing a catalog of ratings, the network including a plurality of interconnected computers with access to the objects, the method comprising:

- a) running on the site a program which processes objects stored on the site,
5 thereby assembling values found in at least one of the objects for comparison to a list of rating values;
 - b) generating a rating for each object by relating the values found in the object to the list of rating values; and
- aggregating the ratings to generate the catalog of ratings.

10 52. A computer data set containing a computer program which, when run on one or more computer systems, causes the systems to perform the method of claim 51.

53. The method of claim 51 wherein each of the rating values comprises a word.

54. The computer data set of claim 52 further comprising the limitations of claim 53.

55. The method of claim 51 wherein the rating values are supplied by a human.

15 56. The computer data set of claim 52 further comprising the limitations of claim 55.

57. The method of claim 51 wherein the rating values are supplied by a computer.

58. The computer data set of claim 52 further comprising the limitations of claim 57.

59. The method of claim 51 wherein generating ratings of the objects comprises comparing the values found in the object to a list of human input rating values supplied by an
20 owner of the site and stored in a file associated with the site which file is read by said program.

60. The computer data set of claim 52 further comprising the limitations of claim 59.

61. The method of claim 59 wherein additional human input rating values are supplied by a host of the site and stored in said file.

62. The computer data set of claim 60 further comprising the limitations of claim 61.

63. The method of claim 51 wherein generating a rating comprises generating a ratings flag when the values found in the object indicate a first rating for the object and at least one of the values from an owner of the site or a host of the site indicates a second rating for the object different than the first rating.

5 64. The computer data set of claim 52 further comprising the limitations of claim 63.

65. The method of claim 51 wherein aggregating ratings includes triggering a human review indicator for review by a human of objects having rating flags to determine the correct ratings for the objects.

66. The computer data set of claim 52 further comprising the limitations of claim 65.

10 67. The method of claim 51 wherein aggregating ratings includes triggering a computer review of objects having rating flags to determine the correct ratings for the objects.

68. The computer data set of claim 52 further comprising the limitations of claim 67.

69. The method of claim 51 wherein the step of processing objects comprises processing meta data for the objects.

15 70. The computer data set of claim 52 further comprising the limitations of claim 69.

71. The method of claim 51 wherein the list of ratings values is stored on the site.

72. The computer data set of claim 52 further comprising the limitations of claim 71.

73. The method of claim 51 wherein the list of ratings values is stored on a second site.

74. The computer data set of claim 52 further comprising the limitations of claim 73.

20 75. A method in a network of computer systems to limit exposure of objects on a computer system to other computer systems on the network, comprising:

- a) assembling a plurality of objects on a computer system connected to a network;
 - b) placing on said computer system an index card file which is not a part of any of said objects and which designates a subset of said objects as accessible to other
- 25 computer systems; and

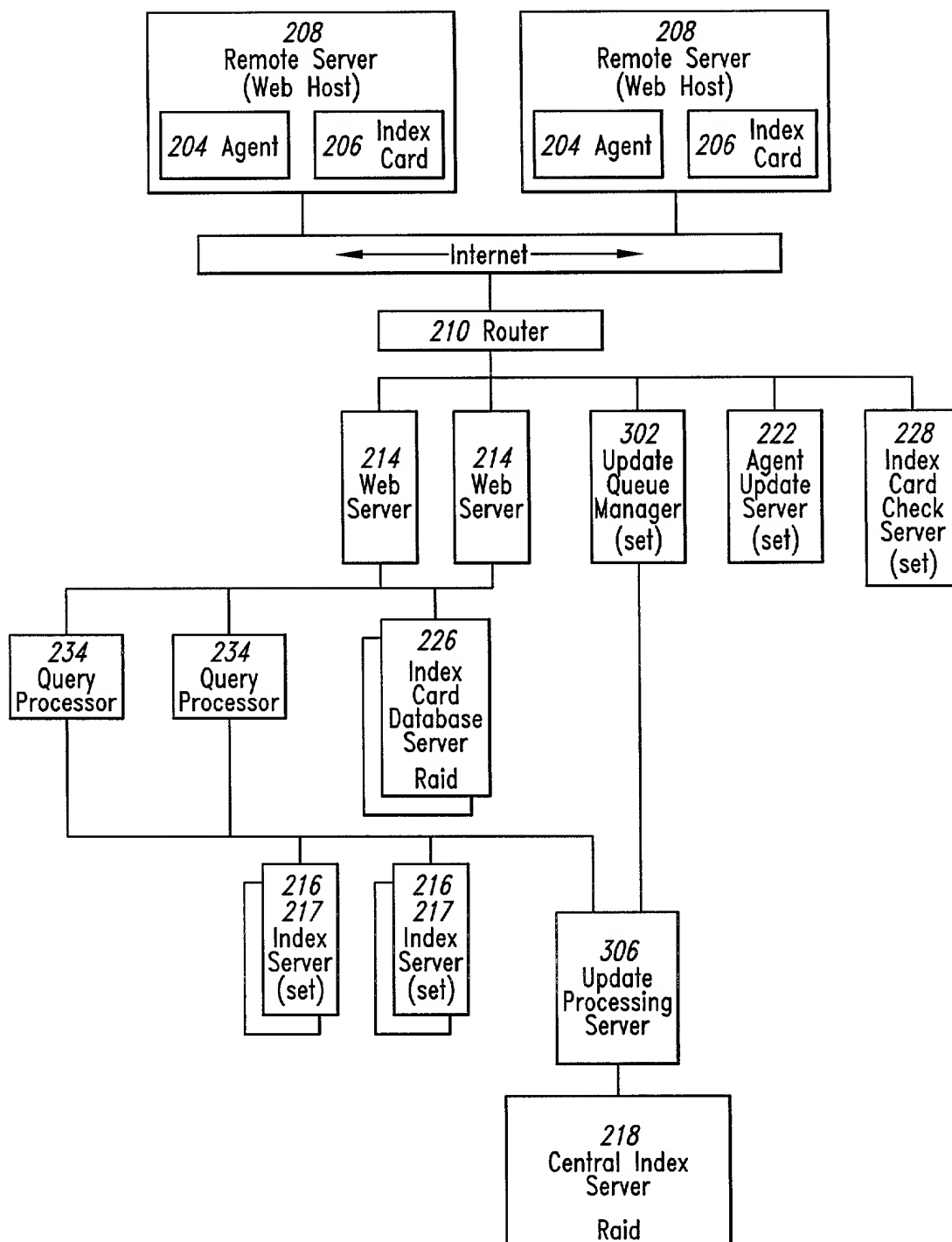
- c) providing to other computer systems information for accessing said objects designated as accessible while retaining as confidential necessary information for accessing other objects on said computer system.

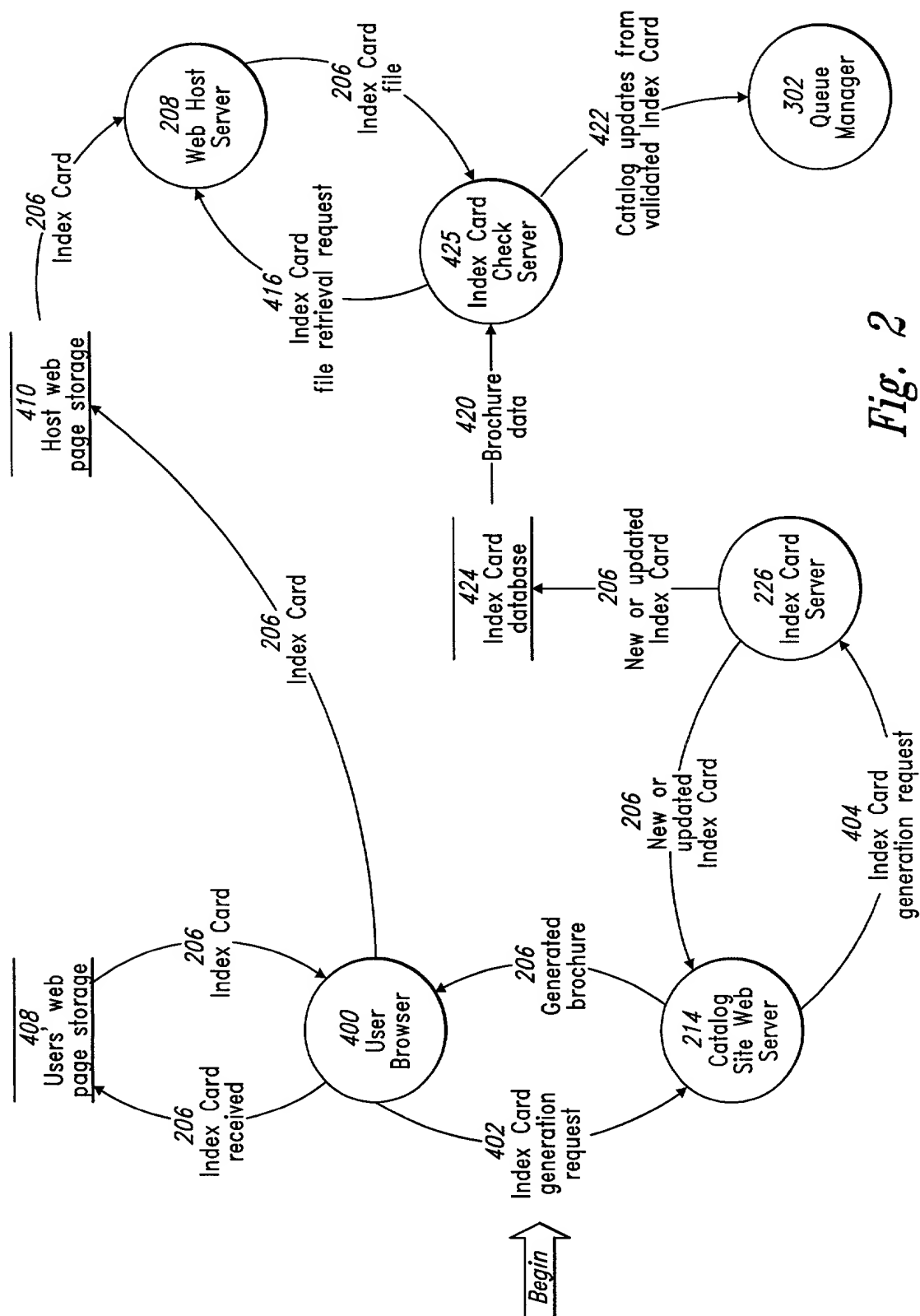
76. A computer data set containing a computer program which, when run on one or more
5 computer systems, causes the systems to perform the method of claim 75.

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
222

Abstract of the Disclosure

Instead of using a central site including spidering software to recursively search all linked web pages and generate a search index of the Internet, index cards files are created and
5 stored on each web site to provide meta data in the form of conceptual or non-keyword information about the site, such as demographics and categorization information. The meta data is periodically collected and aggregated on a central server. A central catalog of object references is compiled on the central site from the meta data reported from each web site. The information from the index cards is utilized in constructing the central catalog so that more
10 accurate search results may be generated for search queries applied to the catalog, so that the search results may be ranked in meaningful ways, and so that web sites may be more accurately rated.

*Fig. 1*



302 Queue Manager 304 Remote Queue Manager 306 Queue database 308 Send queue entries to mirrors 310 Receive queue entries 312 Done receiving 314 Update entries from agents 316 Update entries from confirmed Index Cards 318 425 Brochure Check Server

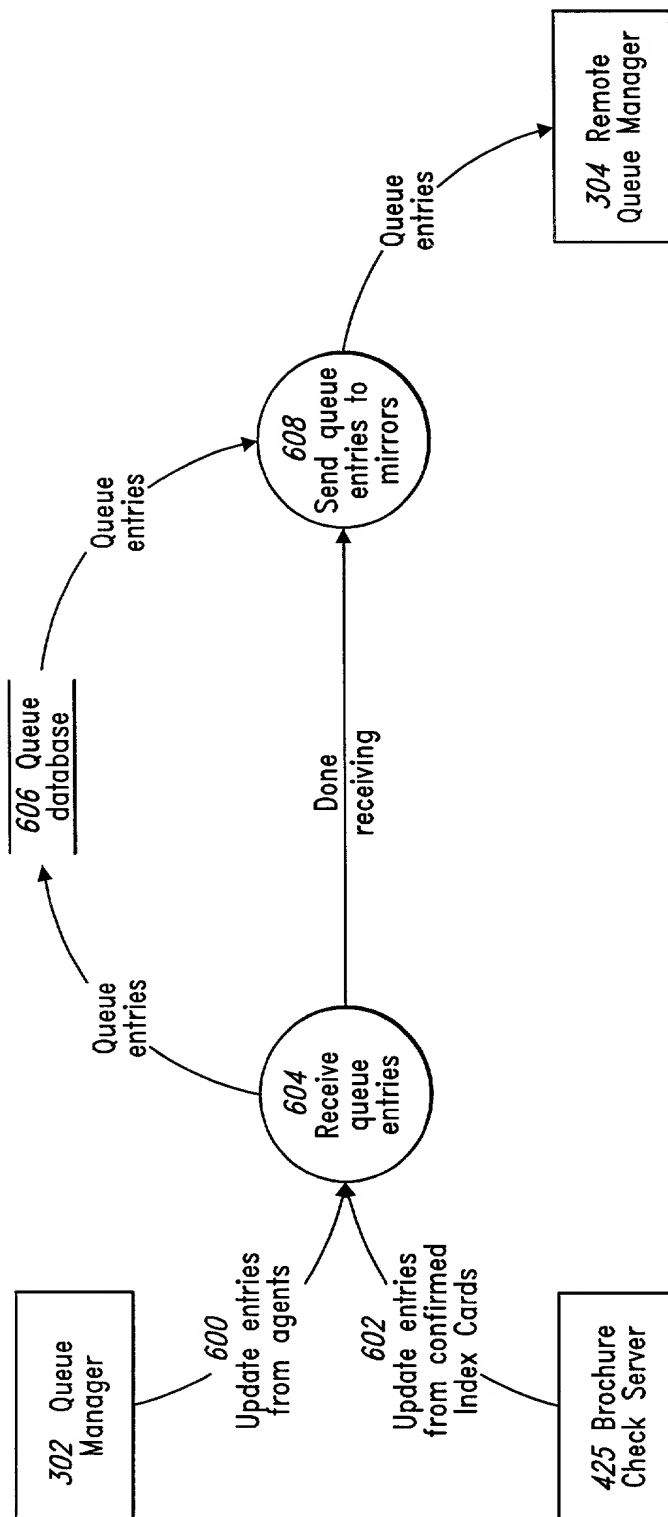


Fig. 3

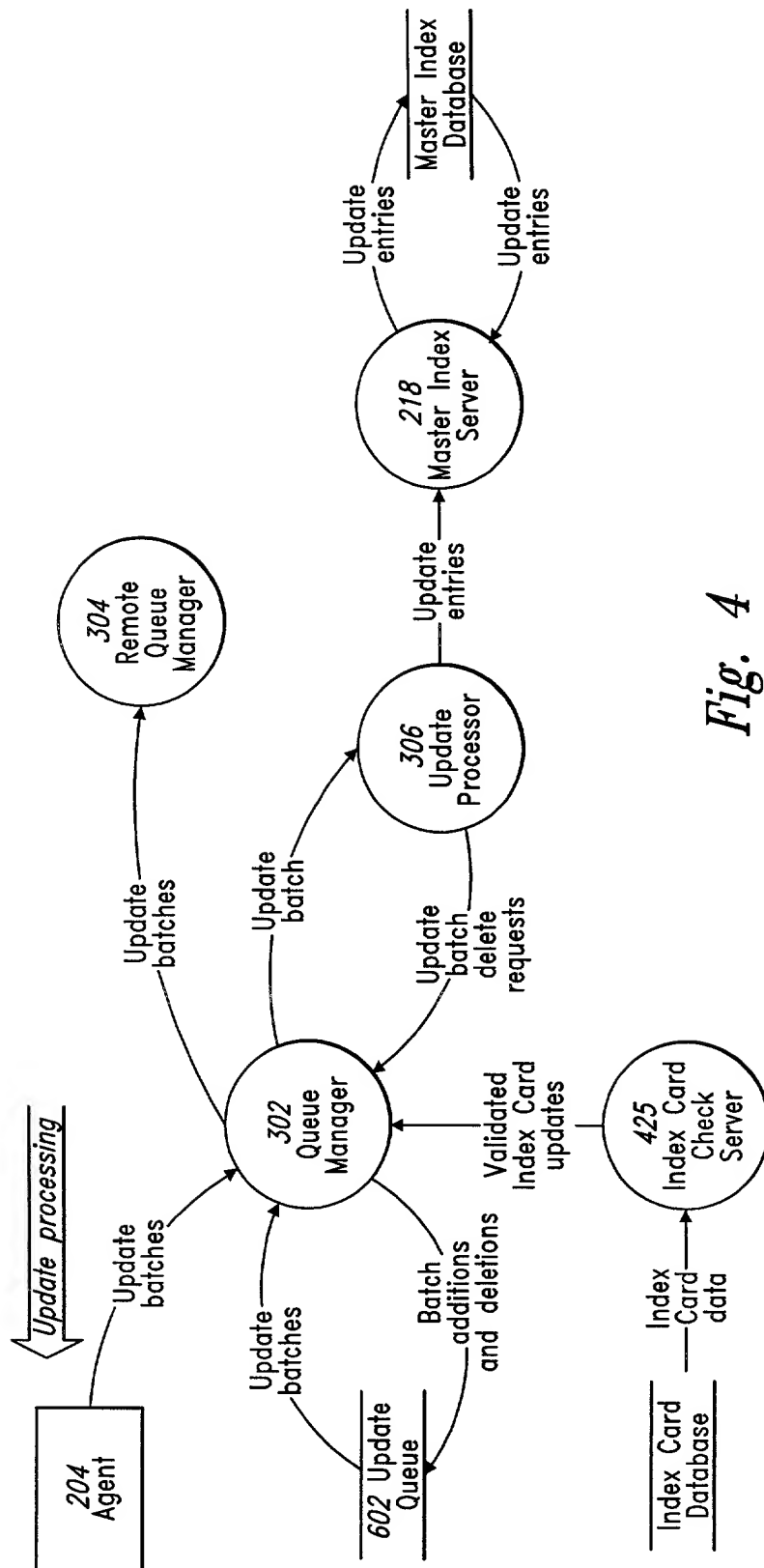


Fig. 4

5/5

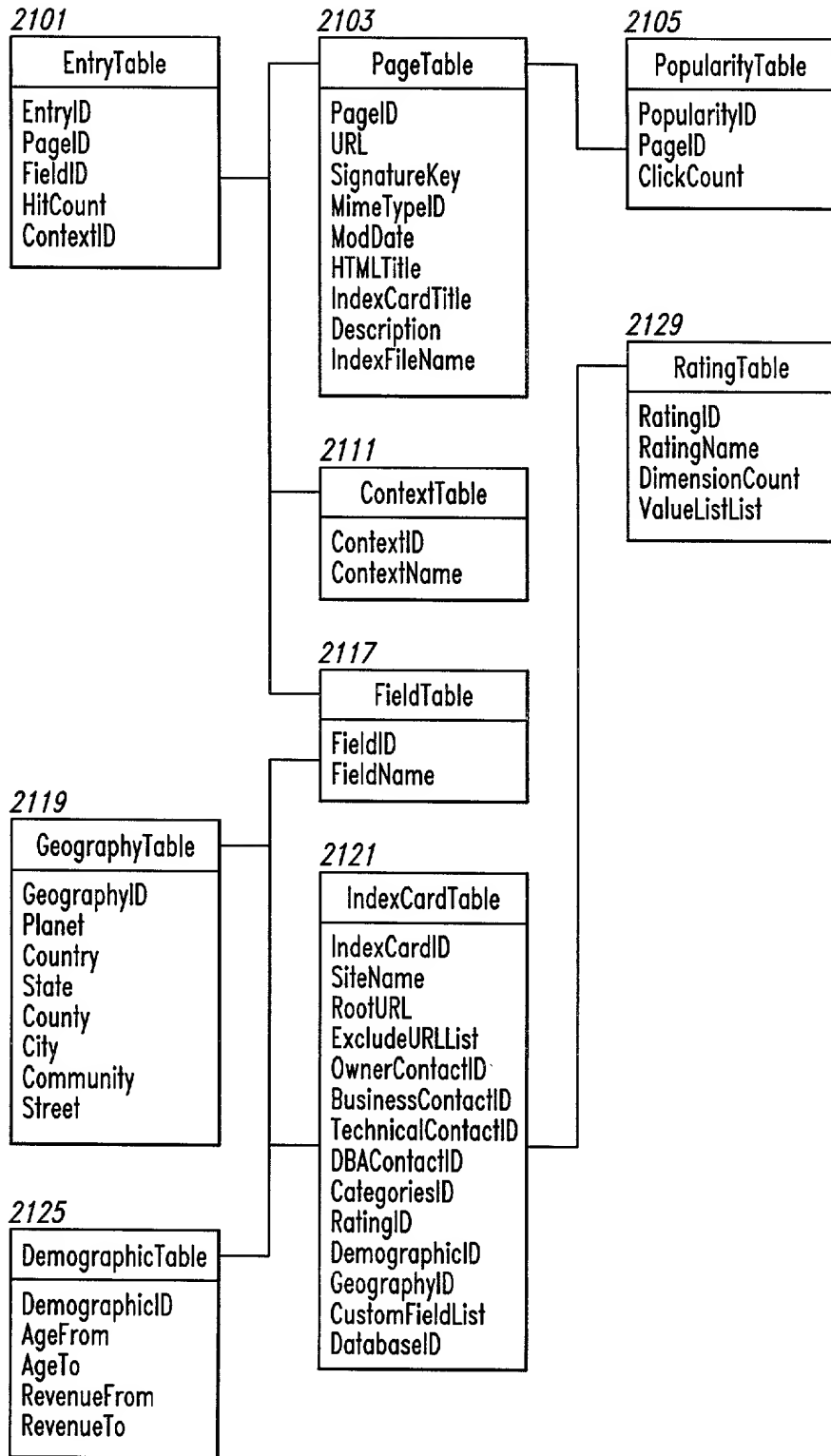


Fig. 5

**DECLARATION AND POWER OF ATTORNEY
IN PATENT APPLICATION**

Attorney Docket No.: 1770-3-3

As a below named inventor, I hereby declare:

My residence, post office address and citizenship are as stated below next to my name.

I believe that I am the original, first and sole inventor (if only one name is listed below) or a joint inventor (if plural inventors are listed below) of the subject matter that is claimed and for which a patent is sought on the invention entitled:

INDEX CARDS ON NETWORK HOSTS FOR SEARCHING, RATING, AND RANKING

the specification of which



is attached hereto.



was filed on _____ as U.S. Application Serial No.
(or PCT International Application No.) _____

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b), of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT international application designating at least one country other than the United States listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s):

<u>Number</u>	<u>Country</u>	<u>Date Filed</u>	<u>Priority Claimed</u>
_____	_____	_____ Day/Mo/Year	_____ <input type="checkbox"/> Yes <input type="checkbox"/> No
_____	_____	_____ Day/Mo/Year	_____ <input type="checkbox"/> Yes <input type="checkbox"/> No

I hereby claim the benefit under Title 35, United States Code Section 119(e) of any United States provisional application(s) listed below.

<u>Application No.</u>	<u>Filing Date</u>
_____	_____

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s), or Section 365(c) of any PCT international application designating the United States listed below, and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT international application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56, which became available between the filing date of the prior application and the national or PCT international filing date of this application.

<u>Application Number</u>	<u>Filing Date</u>	<u>Status: Patented/ Pending/Abandoned</u>
PCT/US00/28653	October 13, 2000	Pending
US 09/419,405	October 14, 1999	Pending

I hereby appoint the attorneys associated with Customer No. 000996 to prosecute this application and to transact all business in the United States Patent and Trademark Office connected therewith. Address all correspondence and phone calls to:

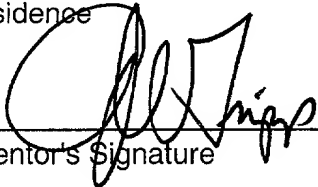
Jeffrey T. Haley
GRAYBEAL JACKSON HALEY LLP
155 - 108th Avenue NE, Suite 350
Bellevue, WA 98004-5901 USA
Telephone (425) 455-5575
Facsimile (425) 455-1046

I hereby further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

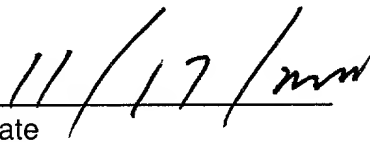
Gary W. Tripp
Full Name of Inventor

United States
Citizenship

9836 Miller Road,
Bainbridge Island, Washington 98110
United States of America
Residence



Inventor's Signature




Date

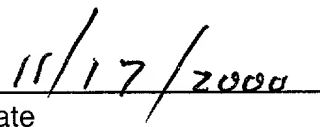
Michael D. Meadway
Full Name of Inventor

United States
Citizenship

18033 - 129th Place Southeast,
Snohomish, Washington, 98290
United States of America
Residence



Inventor's Signature




Date

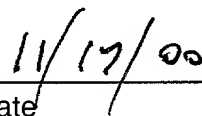
Claude E. Duguay
Full Name of Inventor

United States of America
Citizenship

2835 Boyer Avenue East
Seattle, Washington 98102
United States of America
Residence



Inventor's Signature



Date